

DOCKET NO: 230421US26YA

IN THE UNITED STATES PATENT & TRADEMARK OFFICE

IN RE APPLICATION OF :
ANDREJ S. MITROVIC : EXAMINER: SAXENA, AKASH
SERIAL NO: 10/673,501 :
FILED: SEPTEMBER 30, 2003 : GROUP ART UNIT: 2128
FOR: SYSTEM AND METHOD FOR :
USING FIRST-PRINCIPLES SIMULATION
TO CHARACTERIZE A
SEMICONDUCTOR MANUFACTURING
PROCESS

REPLY BRIEF UNDER 37 CFR 41.41

COMMISSIONER FOR PATENTS
ALEXANDRIA, VIRGINIA 22313

SIR:

This is a Reply Brief to the Examiner's Answer dated November 14, 2008.

TABLE OF CONTENTS

I.	Status of Claims	1
II.	Grounds of Rejection for Review	2
III.	ARGUMENTS.....	3
IV.	Conclusion	16

I. Status of Claims

Claims 1-44 and 48-50 are pending and appealed. Claims 45-47 and 51 are canceled.

Claims 1-44 and 48-50 stand rejected under 35 U.S.C. § 103(a) as being unpatentable over U.S. Patent No. 6,802,045 issued to Sonderman et al., in view of IEEE article “Mathematic-physical engine: parallel processing for modeling and simulation of physical phenomena” by Jain et al.

II. Grounds of Rejection for Review

Whether the rejection of Claims 1-44 and 48-50 under 35 U.S.C. § 103(a) as being unpatentable over Sonderman et al in view of Jain et al should be reversed.

III. ARGUMENTS

Claim 1 defines a method of controlling a process performed by a semiconductor processing tool, including:

- 1) inputting process data relating to an actual process being performed by the semiconductor processing tool,
- 2) inputting a first principles physical model including a set of computer-encoded differential equations, the first principles physical model describing at least one of a basic physical or chemical attribute of the semiconductor processing tool,
- 3) performing first principles simulation for the actual process being performed ***during performance of the actual process*** using the physical model to provide a first principles simulation result in accordance with the process data relating to the actual process being performed in order to simulate the actual process being performed, ***said first principles simulation result being produced in a time frame shorter in time than the actual process being performed***, and
- 4) using the simulation result as part of a data set that characterizes the actual process being performed by the semiconductor processing tool..

The claim defines clearly a process where data input from an actual process being performed is used for producing a first principles simulation result, produced for the actual process being performed during performance of the actual process. The result obtained is produced in a time frame shorter in time than the actual process being performed. The result obtained is then used as part of a data set that characterizes the actual process being performed by the semiconductor processing tool.

The Examiner and the Appellant continue to disagree as to whether Sonderman et al teach performing a first principles simulation for an actual process being performed to provide a first principles simulation result in order to simulate the actual process being performed.

In particular, the Examiner appears to be extensively relying on Figure 1 of Sonderman et al. The Examiner characterized Sonderman et al by stating on page 12 of the Examiner's Answer that:

Here examiner would like to emphasize that new control inputs X_{Ti} are generated to control the "subsequent [part of the] process" (applicable to same processes as inputs are pertinent to same process only - e.g. metal deposition on substrate where inputs may specify to deposit more metal) on the silicon wafer S_i .

Arguendo, if Sonderman was intending to use the inputs for the next wafer, he would have stated for silicon wafer S_{i+1} , with emphasis on $i+1$. Therefore as the limitation **"for the actual process being performed during performance of the actual process"**, is performed during the processing of silicon wafer. (See Sonderman Fig. 1). [Reproduced with emphasis added by the Examiner]

Appellant would like to point out that the Examiner's characterization of Sonderman et al given above does not accurately depict what Sonderman et al disclose. Firstly, Sonderman et al disclose "a subsequent process" not "the subsequent part of the process" as the Examiner has stated. The characterization of Sonderman et al has a significantly different meaning than the plain words in Sonderman et al. Secondly, the parenthetical statements in the Examiner's characterization are mere speculation, and do not represent disclosures in Sonderman et al.

Furthermore, the Examiner in response to Appellant's prior arguments that the simulation results in Sonderman et al are based on historical data characterized Sonderman et al by stating on pages 13-14 of the Examiner's Answer that:

As clearly seen, the control parameter from the process control environment are first feed into the simulation environment, and then the post simulation output, having modified control parameters, is used to control the process control environment. Thus, the results are not based on the historical data for another run, but are from the same run, where the input from the process control environment provided to simulation tool is used to generate modified control parameter for the process control environment. [Reproduced with emphasis added by the Examiner]

Figure 1 of Sonderman et al is reproduced below.

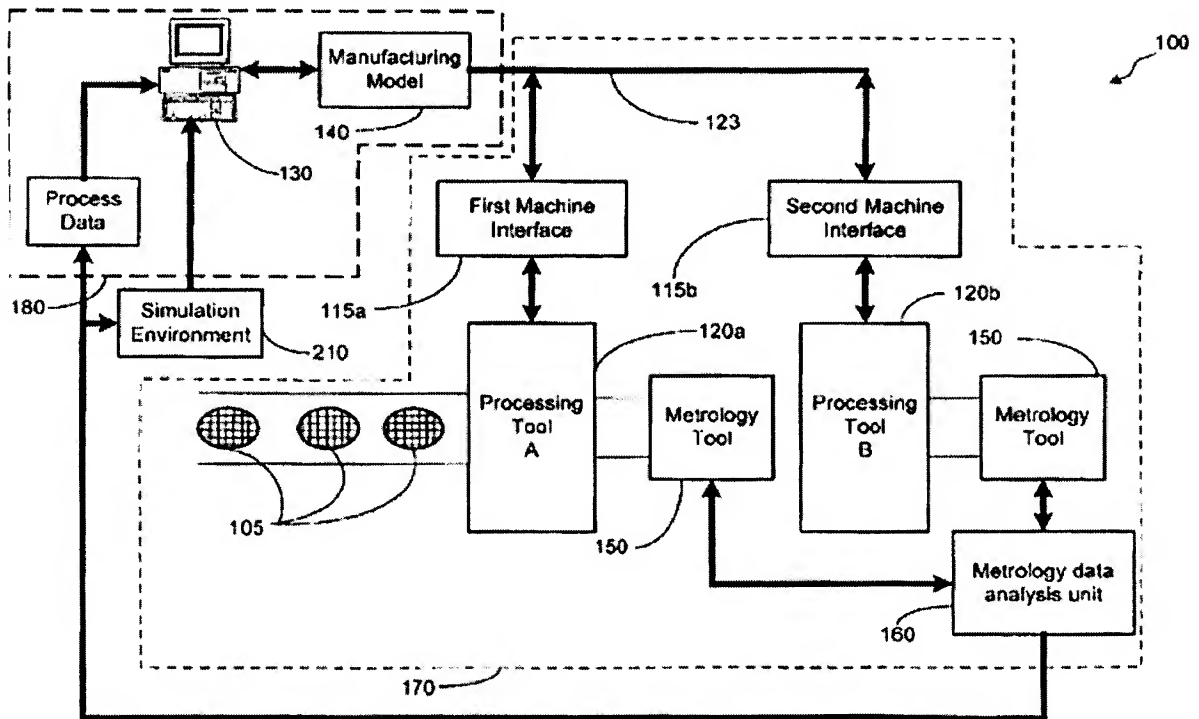


FIGURE 1

Appellant respectfully points out that, at col. 4, line 48, to col. 5, line 10, Sonderman et al specifically states:

Referring now to FIGS. 1 and 2 simultaneously, one embodiment of an interaction between a process control environment 180, a manufacturing/processing environment 170, and a simulation environment 210 is illustrated. In one embodiment the process control environment 180 receives input data from the simulation environment 210, which is then used to control the operation of the manufacturing environment 170. The integration of the simulation environment 210 and the process control environment 180 into the manufacturing environment 170 facilitates more accurate control of the processing of semiconductor wafers. The simulation environment 210 allows for testing various manufacturing factors in order to study and evaluate the interaction between the manufacturing factors. This evaluation can be used by the system 100 to prompt the process control environment 180 to invoke more accurate process control.

Furthermore, the simulation environment 210 can be used for ***feedback modification*** of control parameters invoked by the process control environment 180. For example, the manufacturing environment 170 ***can send metrology data results*** into the simulation environment 210. The simulation environment 210 can then use the metrology data results and perform various tests and calculations to provide more accurate, modified control parameters to the process control environment 180. A feedback loop is then completed when the process control environment 180 sends the modified or adjusted process control parameters to the manufacturing environment 170 for further processing of semiconductor wafers.

Appellants point out that there is **no** disclosure here of the claimed performing a first principles simulation for ***the actual process being performed*** to provide a first principles simulation result in order to simulate the actual process being performed.

Reiterating in part, Claim 1 defines:

inputting process data relating to an actual process being performed by the semiconductor processing tool;

performing a first principles simulation for the actual process being performed to provide a first principles simulation result in order to simulate the actual process being performed.

Thus, for Sonderman et al to meet these claim elements, Sonderman et al would have to show input data from the process being directly used by the computer system 130 to compute a simulation result. Yet, Figure 1 of Sonderman et al shows the **only** input to the computer system 130 is by way of metrology data.

Metrology data is data taken on samples after a process has been performed. See Appellant's specification numbered paragraph [0106] for a listing of metrology data taken on a sample from "a process performed by semiconducting processing tool 102." Indeed, the second part of the noted above quote from Sonderman et al, states with regard to metrology data and feedback control that:

The simulation environment 210 can then use ***the metrology data results*** and perform various tests and calculations to provide more accurate,

modified control parameters to the process control environment 180. A ***feedback loop*** is then completed when the process control environment 180 sends the ***modified or adjusted process control parameters*** to the manufacturing environment 170 ***for further processing*** of semiconductor wafers.

If Sondermann et al were able to obtain metrology data on the actual process being performed, Sondermann et al would not have to restrict the use of metrology data to feedback control, but rather could use the metrology data for feedforward control.

More significantly, Appellant's position on this matter is corroborated by col. 4, lines 31-47, of Sonderman et al which specifically states:

One or more of the semiconductor wafers 105 ***that are processed*** by the processing tools 120a, 120b can also be sent to a metrology tool 150 for acquisition of metrology data. The metrology tool 150 can be a scatterometry data acquisition tool, an overlay-error measurement tool, a critical dimension measurement tool, and the like. In one embodiment, ***one or more processed semiconductor wafers are examined*** by the metrology tool 150. Data from the metrology tool 150 is collected by a metrology data analyzer unit 160. The metrology data analyzer unit 160 organizes, analyses, and correlates scatterometry metrology data acquired by the metrology tool 150, to particular semiconductor wafers 105 that were examined. The metrology data analyzer unit 160 can be a software unit, a hardware unit, or a firmware unit. In one embodiment, the metrology data analyzer unit 160 is integrated into the computer system 130.

Appellant's position on this matter is also corroborated by col. 9, lines 46-51, of Sonderman et al which specifically states:

The system 100 ***then*** optimizes the simulation (described above) to find more optimal process target (T_i) for each silicon wafer, ***S_i to be processed***. These target values are then used to generate ***new control inputs, X_{T_i}*** on the line 805 to control ***a subsequent process of a silicon wafer S_i*** . [Emphasis added]

The plain reading of this section of Sonderman et al is that the system 100 ***then*** (e.g., at time T_1) optimizes the simulation for each silicon wafer, ***S_i to be processed*** (e.g., later at time T_2). In other words, the simulation results of Sonderman et al produce a new control input for each silicon wafer ***to be processed***.

Thus, Appellant respectfully submits that Sonderman et al teach performing a simulation result for a process to be performed *before* performance of the actual process, and do not teach the claimed performing first principles simulation *for the actual process being performed during performance of the actual process*.¹

Other sections of Sonderman et al support Appellant's position on this matter that the simulation results in Sonderman et al are made prior to controlling a subsequent process. For instance, Figure 4 of Sonderman et al (reproduced below) shows that the simulation results are produced *ahead of performing a process* and thus are based on historical data.

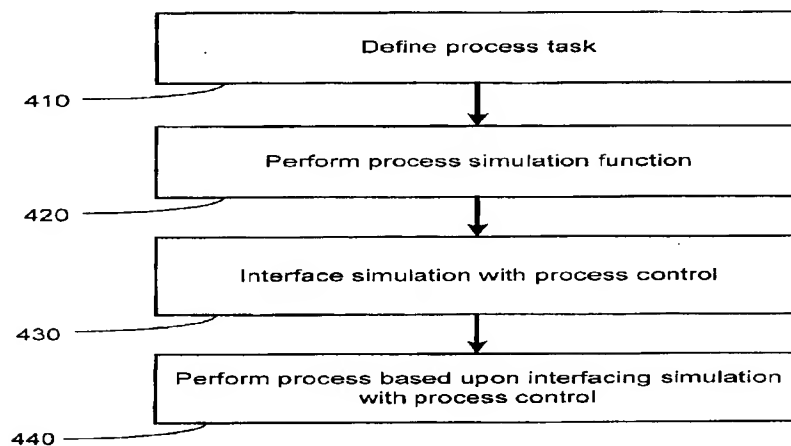


FIGURE 4

With reference to Figure 4, Sonderman et al disclose at col. 6, lines 24-47:

Turning now to FIG. 4, a flow chart representation of the methods in accordance with the present invention is illustrated. In one embodiment, *the system 100 defines a process task that is to be performed (block 410)*. The process task may be a photolithography process, an etching process, and the like. *The system 100 then performs a process simulation function (block 420)*. A more detailed description of the process simulation function

¹ Appellant also points out that Sonderman et al do not disclose or suggest a first principles simulation.

described in block 420, is illustrated below. In one embodiment, a simulation data set results from the execution of the process simulation function.

Once the system 100 performs the process simulation function, the system 100 performs an interfacing function, which facilitates interfacing of the simulation data with the process control environment 180 (block 430). The process control environment 180 can utilize the simulation data in order to modify or define manufacturing control parameters that control the actual processing steps performed by the system 100. ***Once the system 100 interfaces the simulation data with the process control environment 180, the system 100 then performs a manufacturing process*** based upon the manufacturing parameters defined by the process control environment 180 (block 440). [Emphasis added]

Hence, the process flow in Sonderman et al is straightforward:

- 1) define a process to be modeled,
- 2) model the simulation result,
- 3) interface simulation result to processor, and then
- 4) run the process under control based on the pre-existing simulation result.

Accordingly, Appellant respectfully submits that Sonderman et al do not disclose and indeed ***teach away*** from the present invention where data input from an actual process being performed is used for producing a first principles simulation result, which is produced for the actual process being performed during performance of the actual process in a time frame shorter in time than the actual process being performed.

The deficiencies in Sonderman et al for teaching 1) performing first principles simulation for the actual process being performed ***during performance of the actual process*** using the physical model to provide a first principles simulation result in accordance with the process data relating to the actual process being performed in order to simulate the actual process being performed, and 2 that the ***first principles simulation result being produced in a time frame shorter in time than the actual process being performed*** are not overcome by Jain et al. The Office Action in rejecting the present claims supplements the teachings of

Sonderman et al with the teachings of Jain et al for their teaching of computer encoded differential equations in a mathematical physical engine (MPE) which can be applied to wafer processing. See Office Action, page 16. Jain et al describe at pages 372-373 that:

We **propose** a wafer scale implementation of the MPE. The starting point would be a dedicated processing cell, optimized specifically for the PDE arithmetic and data routing. Because of the relative simplicity of the cell, it is expected that extremely large arrays (8x8 to 32x32) **could be** successfully processed on a single piece of silicon using Wafer Scale Integration techniques. In fact, we have already laid the foundation for the development of such a processing cell. Our Universal Multiply-Subtract-Add [11] **could be** adapted for this first cell design. Similarly, our nonlinear coprocessor cell [12]-[14] **might be used** in conjunction with the UMSA to provide advanced mathematical functions. As suggested in Fig. 2, there would be **courtyards of processors**, each with two interconnection networks and two memory banks. 2-D, 3-D, and 4-D problems could then be mapped for parallel computations. Since inter-processor delays are very small (say a few ns), extremely high speeds could be achieved. This, together with the high degree of parallelism, would result also in high throughput. We **envision** 100 to 1000 processors (on one wafer) forming a wafer scale MPE. At a clock frequency of 50 MHz, a single wafer could achieve up to 20 GFLOPs performance. With our nonlinear coprocessor added, each instruction could equate to multiple floating point operations.

Furthermore, because of the extendible architecture, several wafers **could be** interconnected as shown in Fig. 5 to construct a "stacked" MPE wafer system (SMPE). Note that no vertical interconnects within the stack of wafers are expected. Tens to hundreds of GFLOPs performance in a volume the size of a desk-top computer [15] **could** thus be achieved. However, **these predictions** ignore the likely technical advances in the next five years; a tenfold further increase in performance **might be achievable**. [Emphasis Added]

Thus, as emphasized above, the proposed development work in Jain requires the development of **futuristic** computational equipment which one of ordinary skill in the art would be reluctant to implement or utilize for the rigorous standards needed in semiconductor manufacturing.

Appellant's position of the deficiencies of Sonderman et al and Jain et al is corroborated by statements in Appellant's specification. The Examiner takes a position in the Examiner's Answer (see page 17) that there is an inconsistency with regard to Appellant's

specification at numbered paragraphs [004] and [005]. Reproduced below are the entirety of numbered paragraphs [004] and [005] from the specification:

These industry and manufacturing challenges have led to interest in more use of computer based modeling and simulation in the semiconductor manufacturing industry. Computer-based modeling and simulation are increasingly being used for prediction of tool performance during the semiconductor manufacturing tool design process. The use of modeling allows the reduction of both cost and time involved in the tool development cycle. Modeling in many disciplines, such as stress, thermal, magnetics, etc., has reached a level of maturity where it can be trusted to provide accurate answers to design questions. Moreover, computer power has been increasing rapidly along with the development of new solution algorithms, both of which resulted in reduction of time required to obtain a simulation result. ***Indeed, the present inventors have recognized that a large number of simulations typically done in the tool design stage can presently be run in times comparable to wafer or wafer cassette processing times.*** These trends have led to the suggestion that simulation capability typically used only for tool design can be implemented directly on the tool itself to aid in various processes performed by the tool. For example, the 2001 International Technology Roadmap for Semiconductors identifies issues impeding the development of on-tool integrated simulation capability as an enabling technology for manufacturing very small features in future semiconductor devices.

Indeed, the failure of industry to implement on-tool simulation to facilitate tool processes is primarily due to the need for computational resources capable of performing the simulations in a reasonable time. Specifically, the processor capabilities currently dedicated to semiconductor manufacturing tools are typically limited to diagnostic and control functions, and therefore could only perform relatively simple simulations. Thus, the semiconductor manufacturing industry has perceived a need to provide powerful dedicated computers in order to realize meaningful on-tool simulation capabilities. However, dedication of such a computer to the semiconductor processing tool results in wasted computational resources when the tool runs processes that use simple simulations, or no simulations at all. This inefficient use of an expensive computational resource has been a major impediment to implementation of simulation capabilities on semiconductor processing tools.

Thus, this part of the specification specifies that the “present inventors” ***not prior work by others*** have recognized that a large number of simulations typically done in the tool design stage can presently be run in times comparable to wafer or wafer cassette processing

times. Furthermore, this description does not indicate that a first principles simulation result could be produced in a time frame shorter in time than the actual process being performed, as claimed.

Indeed, the “failure of industry” noted in Appellant’s specification emphasized above is evidenced by Tan et al and Kee et al of record.

Tan et al teach the use on an existing process model for feedback or feed forward processing. In feedback control, by definition, the results of a process step are provided to subsequent wafer. In feed forward control, the results of a prior process step are used to adjust a subsequent process being run of the wafer. Tan et al describe:

The illustrative APC Framework 200 includes a process model 202 that receives ***feed-forward and feed-back data*** and calculates a processing parameter. The illustrative portion of the APC Framework 200 includes two measurement devices, in particular a pre-process metrology machine 204 and a post-processing metrology machine 206. The pre-process metrology machine 204 performs a measurement on a material prior to processing in a processing machine 208 and sends the measurement, as feed-forward data, to the process model 202. The processing machine 208 sends processed material to the post-processing metrology machine 206 ***to measure post-process data which is sent to the process model 202 as feedback data.***

Referring to FIG. 4, a schematic block diagram shows material flow of a processing step 400 of a semiconductor manufacturing process from a process engineer perspective. An APC plan 402 retrieves a process model from the data store 306, then executes a parameter calculation algorithm 404. The APC plan 402 gives the calculated parameters to a machine 406 and directs the machine 406 to execute the process. The machine 406 issues a signal to the APC plan 402 ***when the process execution is complete.*** The APC plan 402 sends the calculated parameters to the data history store 310 of the historical database 312.

Referring to FIG. 5, a schematic block diagram shows material flow of a post-process measurement step 500 of a semiconductor manufacturing process from a process engineer perspective. An APC plan 502 sends a message to a machine 504 instructing the machine 504 to measure a post-processed material. The machine 504 sends measurement data to the APC plan 502. The APC plan 502 retrieves an old process model from the data store 306. The APC plan 502 executes a model update algorithm 506. The APC plan 502 ***stores an updated model in the data store 306 for usage in the***

processing step 400. The APC plan 502 sends new model data to the data history store 310 of the historical database 312. [Emphasis added.]

Thus, Tan et al use post-process data to update a model for a subsequent process step.

Kee et al deal with the process control of a Rapid Thermal Processing (RTP) tool and do **not** use real time modeling. RTP tools are tools used in semiconductor manufacturing.

Kee et al in detail disclose that:

The modeling apparatus 101 of the instant invention may also be used to perform an inverse analysis to establish the boundary conditions or parameter values required to achieve a certain function of the thermal system. This allows the apparatus to be used to establish the appropriate process parameters and boundary conditions for the thermal system modeled. In accordance with the instant invention, the inverse analysis can be directly carried out by the modeling apparatus ***rather than using the conventional approach, which merely solves the direct problem repeatedly, in a lengthy and costly iterative process,*** to determine appropriate input parameters to achieve a desired result. In other words, in accordance with the instant invention, ***once a particular thermal process is modeled for a particular set of control parameters,*** the device may then be used to automatically obtain the necessary control parameters to achieve a desired result by providing the modeling apparatus with parameters corresponding to the desired result.

To carry out the inverse analysis, the modeling apparatus 101 includes an inverse parameter input section 104 also connected to input device 103. A user inputs into the modeling apparatus 101 parameters corresponding to desired results, e.g., desired temperature characteristics of the system, which are stored in memory 108. The processing unit 110, under control of modeling program 111, ***uses the previously generated model*** of the thermal system and the parameters held in memory 108 and derives or predicts particular control parameters to meet the constraints entered through the inverse parameter input section 104. This process is more fully described below in connection with the examples provided.² [Emphasis added.]

Hence, Kee et al explicitly disclose that a ***previously generated*** model of the thermal system is used to design and control the thermal system. Kee et al exemplify the difficulties of a “conventional approach” which solves spectral radiation transport equations through “a

² Kee et al, col. 4, lines 21-50.

lengthy and costly iterative process.” These problems forced Kee et al to use *pre-generated model results* for control of a RTP process.

Hence, Kee et al. further Appellant’s position on non-obviousness by illustrating the *technological difficulty* and *failure of others* in producing a first principle simulation result in a time frame shorter in time than the actual process being performed.

Recently published guidelines for the Patent and Trademark Office, published in Federal Register Vol. 72, No. 195, on Wednesday October 10, 2007 entitled: “Examination Guidelines for Determining Obviousness under 35 U.S.C. 103 in View of the Supreme Court Decision in KSR International v. Teleflex Inc,” indicate that:

Office personnel should consider all rebuttal evidence that is timely presented by the Applicants when reevaluating any obviousness determination. Rebuttal evidence may include evidence of “secondary considerations,” such as “commercial success, long felt but unsolved needs, [and] *failure of others*”, and may also include evidence of unexpected results. Office personnel must articulate findings of fact that support the rationale relied upon in an obviousness rejection. As a result, Applicants are likely to submit evidence to rebut the fact finding made by Office personnel. For example, in the case of a claim to a combination, Applicants may submit evidence or argument to demonstrate that:

- (1) one of ordinary skill in the art could not have combined the claimed elements by known methods (e.g., *due to technological difficulties*);
- (2) the elements in combination do not merely perform the function that each element performs separately; or
- (3) the results of the claimed combination were *unexpected*.

Once the Applicant has presented rebuttal evidence, Office personnel should reconsider any initial obviousness determination in view of the entire record. All the rejections of record and proposed rejections and their bases should be reviewed to confirm the continued viability. The Office action should clearly communicate the Office’s findings and conclusions, articulating how the conclusions are supported by the findings. [Emphasis Added.]

M.P.E.P. § 2143.01(II) indicates that all teachings in the prior art must be considered.

M.P.E.P. 2141.03 indicates that the examiner must ascertain what would have been obvious to one of ordinary skill in the art at the time of the invention. Hence, for all these legal

considerations, Appellant's description in the specification of the failure of industry to provide computational resources capable of performing the simulations in a reasonable time, and the failure of Tan et al and Kee et al to produce a first principles simulation result in a time frame shorter in time than the actual process being performed are **all** presented here as rebuttal evidence for the non-obviousness of the claims.

For all these reasons, Appellant submits that independent Claims 1, 23, and 48 patentably define over Sonderman et al and Jain et al and Tan et al.

Hence, the 35 U.S.C. § 103(a) rejection of Claims 1-44 and 48 as being unpatentable over Sonderman et al in view of Jain et al should be reversed.

Claim 49 (and similarly Claim 50) defines that the performing a first principles simulation includes providing for the first principles simulation a reuse of known solutions as initial conditions for the first principles simulation. The Office Action notes that "Jain teaches use of Navier Stokes and other known simulation solutions" and cites pp. 367-368 of Jain et al. However, the Navier Stokes equation on page 367 of Jain et al is a fluid flow equation which needs boundary conditions and which need s to be solved in order to produce a solution. In effect, the Navier Stokes equation on page 367 of Jain et al is a **problem** to be solved, not a solution. The Navier Stokes equation on page 367 of Jain et al does **not** represent a solution, much less the reuse of known solutions as initial conditions for the first principles simulation. Appellant's inspection of the remainder of Jain et al finds no disclosure of the reuse of known solutions as initial conditions for the first principles simulation.

Hence, for this additional reason (besides their dependence from allowable claims), the 35 U.S.C. § 103(a) rejection of Claims 49 and 50 as being unpatentable over Sonderman

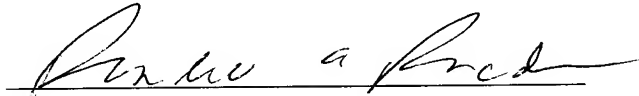
et al in view of Jain et al should be reversed.

IV. Conclusion

Appellant request on the basis of the arguments presented above that the outstanding grounds for the rejection be reversed. Appellant submits that the application is in condition for allowance.

Respectfully submitted,

OBLON, SPIVAK, McCLELLAND,
MAIER & NEUSTADT, P.C.



Steven P. Weihrouch
Registration No. 32,829
Attorney of Record

Ronald A. Rudder, Ph. D.
Registration No. 45,618

Customer Number

22850

Tel: (703) 413-3000
Fax: (703) 413 -2220
(OSMMN 06/04)